



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica
Relazione

**Edizione digitale (diplomatica-interpretativa) dei
documenti della Pace di Dante conservati presso
l'Archivio Storico di La Spezia**

Candidato: *William Brai*

Relatore: *Prof. Angelo Mario Del Grosso*

Correlatore: *Prof.ssa. Enrica Salvatori*

Anno Accademico 2022-2023

Indice

Introduzione	3
1. La Pace di Dante	4
1.1 Origine e descrizione	5
1.2. Composizione e conservazione	6
1.3 Le edizioni	7
1.4 Archivio della Spezia	8
2. La codifica digitale	9
2.1 I linguaggi di markup	9
2.1.1 XML	10
2.1.2 TEI	12
2.2 Edizioni digitali	14
2.2.1 Tipologie	15
3 EVT-Edition Visualization Technology	16
3.1 Funzionalità	17
3.2 Seconda Versione	17
3.3 Architettura	18
4 Progetto di codifica della Pace di Dante	19
4.1 Pianificazione concettuale	19
4.2 Digitalizzazione delle immagini e trascrizione dei testi	20
4.3 Criteri di edizione e-markup	20
4.4 Livelli di edizione del testo	25
4.4.1 Edizione diplomatica-interpretativa	25
4.4.2 Edizione critica	30
4.5 Dal processo di codifica alla visualizzazione	33
4.6 Confronto con altri progetti	38
4.7 Possibili sviluppi	39
5 Conclusioni	40
6 Bibliografia	41
7 Sitografia	42

Introduzione

Il progetto¹ per la realizzazione dell'edizione digitale dei documenti appartenenti alla Pace di Dante nasce con lo scopo di raccogliere le importanti testimonianze dantesche conservate in Archivio e di renderle fruibili non solo agli studiosi ma anche ad un pubblico più vasto. Per il raggiungimento di questo obiettivo sono state sfruttate le potenzialità di EVT 2 (Edition Visualization Technology)², uno strumento open source progettato specificatamente per la presentazione web di edizioni digitali. Questo strumento elabora testi codificati secondo le norme del linguaggio XML (Extensible Markup Language)³, nello specifico schema definito dal consorzio della Text Encoding Initiative (TEI), consentendo all'utente di utilizzare strumenti di indagine dinamici e interattivi per la navigazione dei testi. Dopo aver rappresentato il testo con livello di edizione diplomatica-interpretativa sono state implementate alcune funzionalità che ci consentono di evidenziare ed identificare gli interventi editoriali presenti nell'edizione del 2016 del Codice diplomatico Dantesco.

Il primo capitolo del presente contributo fornisce una panoramica sull'insieme dei documenti riferiti con il nome di Pace di Dante, partendo dalle loro origini sino alla conservazione, illustrandone i dettagli e le varie edizioni critiche. Nello stesso capitolo, si mette a fuoco, inoltre, alcuni aspetti importanti di questo determinato nucleo tematico legati alla collaborazione con l'Archivio di Stato della Spezia e alla storia delle varie pubblicazioni dell'opera documentaria.

Il secondo capitolo affronta il tema della codifica digitale sia dal punto di vista tecnico-teorico sia dal punto di vista pratico. Si fornisce una definizione il più possibile non ambigua, facendo chiarezza sulle differenti tipologie di edizioni digitali e sui corrispondenti vantaggi. Si introducono i concetti chiave appartenenti ai linguaggi di markup⁴, indispensabili per il processo di digitalizzazione e rappresentazione dei testi.

¹ Pagina del repository github: <https://github.com/Williambrai97/Progetto-Finale-Brai-William>

² Pagina di riferimento: <https://github.com/evt-project/evt-viewer/>

³ Sito ufficiale W3C: <https://www.w3.org/XML/>

⁴ "L'espressione markup deriva dalla analogia tra i linguaggi e le annotazioni inserite da autori, curatori editoriali e correttori nei manoscritti e nella bozze di stampa di un testo al fine di indicare correzioni e

Nel terzo capitolo si analizzano le caratteristiche del software EVT, descrivendone le funzionalità, l'architettura e le tecnologie utilizzate per il suo sviluppo.

Nel quarto capitolo si descrive il lavoro svolto durante il progetto di tesi illustrando gli strumenti, i linguaggi e le funzionalità adottate. Vengono approfondite le soluzioni impiegate per soddisfare le esigenze di visualizzazione e descritte le modifiche attuate allo scopo di ampliare e adattare le funzionalità del software EVT in accordo alla composizione dei documenti trattati. Nel capitolo conclusivo, infine, si discute su quanto realizzato, analizzando le soluzioni proposte anche in prospettiva di possibili sviluppi futuri del software EVT nonché di possibili collaborazioni con le istituzioni culturali per la valorizzazione e trasmissione del patrimonio culturale della comunità cittadina.

1. La Pace di Dante

Prima di dare uno sguardo dettagliato alla Pace di Dante - un insieme di documenti del 1306 - è bene illustrare il luogo della sua ultima edizione critica: il Codice diplomatico Dantesco. La menzionata raccolta ha avuto diverse edizioni. La prima del 1940 a cura di Renato Piattoli e l'ultima del 2016 a cura di Teresa De Robertis, Laura Regnicoli, Giuliano Milani e Stefano Zamponi. L'opera, nella sua ultima versione, comprende tutti i documenti di cui si abbia notizia riguardanti Dante e la sua famiglia: dal primo esponente attestato fino alla generazione dei figli. Si tratta in totale di 328 atti prodotti dall'anno 1131 al 1417, che formano il nucleo principale dell'edizione. A questi sono stati aggiunti poi altri 347 documenti, raccolti a vario titolo in quattro appendici, relativi a Pietro, figlio di Dante giudice a Verona, ai nipoti fiorentini di Dante e a quelli veronesi, e a documenti di dubbia pertinenza dantesca, in una silloge che nel suo complesso comprende 675 unità. Si tratta di documenti in senso tecnico, nel significato cioè stabilito dalla disciplina storico-filologica che di essi si occupa, la diplomatica: scritture di negozi giuridici o che sono la memoria di fatti di reali o potenziale rilievo giuridico, redatte nell'osservanza di forme, storicamente determinate, capaci di conferire loro validità legale. Ciò significa che di Dante, dei suoi avi e discendenti, si troveranno gli

trattamenti editoriali, chiamate in inglese mark up". Marco Calvo, Fabio Ciotti, Gino Roncaglia, Marco Zela, *Come funziona World Wide Web, in Internet 2000*

acquisti, le vendite, gli affitti di case o di beni fondiari, i prestiti di denaro, le questioni fiscali e giudiziarie, la nomina di procuratori o tutori legali, i testamenti, le carte prodotte nell'esercizio di una professione e le attestazioni di semplice presenza in qualità di testimone.

Gli atti della pace che Dante, in rappresentanza del marchese Franceschino Malaspina, strinse nel 1306 con Antonio vescovo di Luni, sono integrati da due testi assenti nell'edizione precedente di Piattoli, risalente al 1940. Al di là del contributo che a questa pattuizione può aver dato Dante, questi due nuovi testi consentono di cogliere meglio il carattere squilibrato dell'accordo, del tutto favorevole ai Marchesi. Nel complesso, del resto, il gruppo di documenti definisce con precisione un termine cronologico dal quale è necessario partire per riflettere sulla fase in cui Dante risulta al servizio di questa casata principesca insediata nell'Appennino occidentale.

1.1 Origine e descrizione

Le *tabulae* della Pace di Dante sono state rogate presso Sarzana dal notaio Giovanni Parente di Stupio e mettono in luce fatti e personaggi importanti di un periodo storico in cui era in corso sul territorio di Sarzana una disputa tra potentati, quello dei marchesi Malaspina e quello del vescovo di Luni, con un inevitabile scontro di interessi sia sul piano economico che su quello politico. I documenti testimoniano, infatti, l'attività di mediazione diplomatica di Dante, nel conseguire la pacificazione tra i marchesi Franceschino, Morello e Corrado e il vescovo di Luni, Antonio di Nuvolone da Camilla. La prima tabula contiene il *mandatum*, vale a dire la procura plenipotenziaria attribuita dai Malaspina di Mulazzo all'esule fiorentino. L'atto rappresenta una procura, di carattere plenipotenziario, ovvero generale e speciale, che presuppone un'ampia fiducia della famiglia nei confronti di Dante, consentendogli di procedere senza particolari vincoli materiali e giuridici, ad un accordo di pacificazione con Antonio Nuvolone da Camilla, il vescovo della diocesi di Luni.

Le *tabulae* dalla seconda alla quarta sono invece dedicate all'*instrumentum pacis*, ossia alla pace stipulata tra la famiglia Malaspina e il vescovo di Luni, sancita il 6 ottobre 1306 in Castelnuovo (Magra) e suggellata dal bacio, scambiato in pubblico, tra l'ambasciatore Dante ed il vescovo Antonio Nuvolone da Camilla.

L'accordo consentì ad entrambe le parti di poter esercitare contemporaneamente i corrispettivi diritti feudali sui castelli di Brina e di Bolano, considerati tra i principali oggetti del contendere.

La quinta e la sesta *tabula* trattano della *remissio de condemnationis* e dunque degli accordi reciproci di rinuncia ad ogni rivalsa sia economica che giuridica tra le parti, risolvendo le pendenze giudiziarie ancora in corso tra i due schieramenti e soprattutto le contestazioni sui territori di Arcola, Beverino, Bolano, Calice, Sarzana e Carrara.

Infine è presente un'altra *tabula* che dispone di un atto di liberalità a favore dei frati minori del convento di San Francesco in Sarzana.

Tutti i documenti provano la presenza di Dante Alighieri nel 1306 in Lunigiana e Val di Magra, ospitato dal ramo della famiglia Malaspina, detta dello Spino Secco.

1.2. Composizione e conservazione

Le carte dantesche, al momento della stipulazione, furono scritte su un registro notarile in carta, su quattordici facciate, sul recto e sul verso.

I documenti furono rinvenuti fortunatamente nel 1765 (IV centenario della nascita di Dante), nel corso di ricerche d'archivio commissionate dall'ultimo Malaspina del ramo di Terrarossa, il quale rivendicava alcuni diritti sul feudo di Treschietto.

Recentemente, a seguito dell'acquisizione da parte dell'Archivio di Stato della Spezia, le carte sono state sottoposte ad un accurato restauro che ne ha migliorato ed aumentato la leggibilità, anche grazie alla tecnica digitale. Inoltre, sono state immesse in speciali teche di vetro per garantire un'idonea conservazione e una possibile esibizione al pubblico. Sabato 4 settembre 2021, infatti, tramite un evento targato FAI (Fondo Ambiente Italiano), in onore del settecentesimo anniversario della morte di Dante Alighieri, le *tabulae* della Pace di Dante sono state esposte al pubblico.

Per quanto riguarda lo stato di conservazione delle pergamene è difficile fornire una descrizione unitaria. Alcune sono state profondamente danneggiate dall'umidità con conseguente perdita di porzioni del supporto, mentre altre sono state cucite insieme in passato e per questo riportano tagli che compromettono la leggibilità.

1.3 Le edizioni

Come abbiamo già accennato nel paragrafo precedente, i documenti della Pace di Dante sono stati editi criticamente all'interno di una raccolta che porta il nome di Codice Diplomatico Dantesco. Cesare Balbo, nel 1839, fu il primo a suggerire la necessità di una raccolta di documenti danteschi da porre a corredo e fondamento di una futura biografia del poeta, e ad indicarne il titolo per le future edizioni. Prima che il suo suggerimento venisse raccolto passarono quasi sessant'anni. Il primo Codice Diplomatico uscì nel maggio del 1895 a cura di Guido Biagi e Giuseppe Lando Passerini e sotto gli auspici della Società Dantesca Italiana, col sottotitolo *I Documenti della vita e della famiglia di Dante Alighieri*. Il fascicolo era suddiviso in quattordici dispense che a loro volta contenevano quarantanove documenti danteschi riguardanti alcuni scritti sulla biografia del poeta e sui suoi familiari. In seguito, nel 1940, Renato Piattoli diede alla stampa, sotto gli auspici della Società Dantesca Italiana, il suo nuovo Codice diplomatico dantesco. Il complesso documentario messo insieme da Piattoli è formato da 260 pezzi, 230 costituenti la serie principale e 30 distribuiti in due appendici. Nella prima sono raccolti 28 documenti che attestano l'attività di Pietro di Dante come giudice. Sono invece consegnati nella seconda appendice due atti rogati a Lucca il 21 Ottobre 1308 e a Bologna il 13 Agosto 1327, in cui sono attestati rispettivamente un Giovanni Dante Alighieri da Firenze e un Pietro del fu Dante da Firenze scolare in diritto civile.

Dieci anni dopo l'uscita della prima edizione del Codice Diplomatico Dantesco di Piattoli nel febbraio del 1950, in piena guerra, presso la libreria Gonnelli e in 275 esemplari, uscì la seconda edizione corretta dei pochi errori della prima. Infine, dopo 64 anni, nel 2016 è stata pubblicata l'ultima edizione del Codice Diplomatico Dantesco a cura di Teresa De Robertis, Giuliano Miliani, Laura Regnicola e Stefano Zamponi. La nuova raccolta riprende il lavoro precedentemente attuato da Piattoli, con l'integrazione di alcuni documenti e di note ai testi che intendono implementare l'opera con alcuni minimi ragguagli su tre aspetti essenziali all'intelligenza del documento stesso: il contesto documentario, quello storico e quello biografico.

1.4 Archivio della Spezia

Il rapporto di collaborazione con l'Archivio di Stato della Spezia inizia con l'attività di tirocinio - attuata per tramite del Laboratorio di Cultura Digitale - , durante il quale si è elaborata un'edizione critica della Pace di Dante. Questa cooperazione, inoltre, è stata consolidata dalla volontà di contribuire alla digitalizzazione e alla conservazione del patrimonio documentario, fornendo strumenti per una sua maggior accessibile visualizzazione e incrementando la sua diffusione.

L'articolato complesso di documenti conservati nell'Archivio di Stato della Spezia si distingue sia per quantità in termini di consistenza numerica (10.000 documenti), sia per qualità in termini di antichità, unicità e universalità delle fonti.

La Sezione di Archivio di Stato della Spezia fu istituita per la prima volta il 16 febbraio del 1956, tuttavia, intraprese la sua attività solo durante l'anno seguente.

Negli anni successivi, in virtù delle disposizioni previste dalla cosiddetta legge archivistica, il D.P.R. n. 1409 del 30 settembre 1963, l'Istituto spezzino conseguì la sua autonomia distaccandosi da quello di Genova e diventando Archivio di Stato.

L'Archivio è noto specialmente per la sua collezione di libri, periodici e raccolte.

Alcuni tra i testi che illustrano l'unicità dell'Archivio sono di seguito descritti:

- *Clavis sive Repertorium Discepcionum* risalente al 1648.
- La copia in sei volumi degli *Statutorum Civilium Serenissimae Reipublicae Ianuensis* pubblicata nel 1707.
- gli *Statuta Comunitatis Levanti* edito nel 1773.
- il raro volume *dell'Antica Città di Luni e del suo stato presente: memorie raccolte da Carlo Promis*, aggiuntivi il corpo epigrafico lunense, stampato nel 1838 del quale si possiede anche la ristampa del 1857.
- la *Vita di Andrea Doria* del 1864 ad opera di Francesco Domenico Guerrazzi.

Tra la documentazione di recente acquisizione si segnalano le celebri carte della Pace di Dante, ovvero sette *tabulae* originariamente appartenenti a un cartulare notarile, risalenti all'anno 1306 e versate nel 2005 dall'Archivio Notarile della Spezia.

Il lavoro svolto durante il tirocinio e finalizzato con il progetto finale di tesi ha portato alla digitalizzazione e codifica di alcune pagine della raccolta dantesca, promuovendo una maggiore accessibilità e diffusione all'opera stessa.

2. La codifica digitale

Per rendere disponibile l'immenso patrimonio testuale tramite l'uso di sistemi digitali e computazionali è necessario effettuare una trasposizione dei testi dal loro supporto originario verso il nuovo supporto elettronico.

Dobbiamo a Fabio Ciotti una delle più complete ed esaustive definizioni di codifica informatica dei testi. Secondo lo studioso «per codifica informatica dei testi si intende la rappresentazione formale di un testo ad un qualche livello descrittivo, su di un supporto digitale, in un formato utilizzabile da un elaboratore (Machine Readable Form and Machine Actionable Form) mediante un opportuno linguaggio informatico». ⁵ Ogni testo può quindi essere considerato un'unione di due componenti: quella materiale corrispondente al supporto e alle tracce d'inchiostro e quella astratta coincidente alle sequenze verbali, le quali a loro volta generano una serie di livelli semantici e strutturali.

In merito a questo processo si è espresso uno dei pionieri dell'informatica umanistica, Giuseppe Gigliozzi: «Codificare un testo significa esplorare il codice di partenza. Studiarne regolarità e funzioni. Confrontarle con le costrizioni proprie del codice prescelto e riproporre le relazioni (che sono già in un codice) del testo originale del nuovo codice che, nel nostro caso, dovrà essere utilizzabile per un'elaborazione elettronica. Lavoreremo quindi con due codici: il codice del documento di partenza e il codice che renderà questo documento disponibile per una elaborazione elettronica». ⁶

In altre parole, codificare un testo equivale a trascriverlo in una forma che ne consente l'analisi attraverso mezzi informatici, a renderne esplicita la sua interpretazione, a definire e rispettare uno schema di codifica specifico che stabilisca le caratteristiche del testo che si intende esplicitare e, infine, a utilizzare un linguaggio di codifica formale per marcare le caratteristiche del testo.

⁵ Cfr. Fabio Ciotti, *Cosa è la codifica informatica dei testi*, in AA.VV., *Atti del Convegno Umanesimo & Informatica* (Trento 24-25 maggio 1996), a cura di Daniela Gruber e Patrick Pauletto.

⁶ Cfr. Giuseppe Gigliozzi, *Il testo e il computer. Manuale di informatica per studi letterari*, Milano, Bruno Mondadori, 1997, p. 56.

2.1 I linguaggi di markup

I linguaggi di *markup*⁷ vengono utilizzati per aggiungere a un testo tutte le ulteriori informazioni di carattere tipografico o strutturale.

Grazie a questa tipologia di linguaggi vengono inserite una serie di istruzioni, dette tag, ovvero delle etichette che determinano alcune caratteristiche del testo cui fanno riferimento.

I linguaggi di *markup* sono classificati in due categorie: i linguaggi procedurali e quelli dichiarativi. I primi, conosciuti anche come *specific markup language*, includono istruzioni operative, le quali eseguono procedure computazionali sulle porzioni di testo relazionate ad esse⁸. I secondi, detti anche *generic markup language*, invece, non azionano le procedure informatiche attraverso i loro tag, ma piuttosto specificano e dichiarano la funzione attribuita al blocco di testo cui si riferiscono.

L'SGML⁹ è stato il primo metalinguaggio di markup descrittivo, standardizzato a livello internazionale, progettato con l'intento di favorire lo scambio di documenti indipendentemente dall'hardware e dal sistema operativo utilizzato¹⁰.

2.1.1 XML

SGML pur permettendo di separare il contenuto di un documento dal suo stile, tramite una sintassi flessibile e potente, risulta però eccessivamente complesso richiedendo talvolta lunghi tempi di elaborazione. Per questo motivo nel 1996 nacque, come sottoinsieme di SGML, il metalinguaggio XML (*Extensible Markup Language*), con caratteristiche simili ma di complessità inferiore.

XML è un metalinguaggio di marcatura (*Markup Language*) estensibile (*Extensible*) in quanto permette di creare elementi e strutture personalizzate.

Oltre a consentire l'inserimento di dati strutturati in un file di testo, servendosi dell'insieme di regole, di linee guida e di convenzioni incluse in esso, riesce, inoltre, a

⁷ In particolare, si classificano come markup language solo quei linguaggi che si basano sul testo, sia per il markup sia per il contenuto.

⁸ Uno dei linguaggi più comuni appartenente a questa categoria è l'RTF (Rich Text Format), il quale sfrutta il tipo di codifica del "tagged", in cui il documento viene codificato come puro testo, mentre degli identificatori (tag) indicano quale tipo di struttura di formattazione applicare al semplice testo.

⁹ Sito di riferimento: <https://www.iso.org/standard/16387.html/>.

¹⁰ Per maggiori informazioni riguardo la storia di SGML si veda: SGML Users' Group, *A Brief History of the Development of SGML*, <<http://www.oasis-open.org/cover/sgmlhist0.html>> 11 giugno 1990.

produrre file facilmente generabili e leggibili, ovviando le problematiche legate all'ambiguità, l'estensibilità e alla supportabilità tra le differenti piattaforme.

In un documento XML le proprietà semantiche vengono assegnate tramite l'inserimento di marcatori detti tag, i quali consentono di effettuare complesse ricerche nel testo¹¹. I dati del documento XML sono organizzati gerarchicamente secondo una struttura ad albero la cui radice, contenente tutti gli altri elementi, è rappresentata dall'elemento principale. XML, allo stesso modo di SGML, sfrutta la DTD (*Document Type Definition*)¹² per definire le regole sintattiche con cui rappresentare la struttura dei dati.¹³ Un aspetto importante di XML è quello legato al concetto di "ben formato". Un'istanza XML viene riconosciuta come ben formata se è presente l'elemento radice e se è corretto l'anidamento e la sintassi degli elementi. L'analisi della validità¹⁴ dei documenti XML è affidata a software detti *parser*. Una volta riconosciuto un file XML come ben formato, è possibile trasformarlo in una tipologia di documento differente e, infine, effettuare una sua visualizzazione attraverso l'HTML¹⁵.

¹¹ La funzione di ricerca è comunemente presente in qualunque editor di testi e da la possibilità, appunto, di individuare nel testo una determinata sequenza di caratteri specificata dall'utente.

¹² "Una DTD mette a disposizione un modello per la marcatura di documenti che indica la presenza, l'ordine e la posizione degli elementi e dei loro attributi in un documento." Cfr. Devan Shepherd, *XML. Guida completa*, trad. it. staff Apogeo, Milano, Apogeo, 2002, p. 54 (ed. orig. *Teach yourself XML in 21 Days second edition*, Sams, 2001).

¹³ Le DTD del XML hanno meno elementi ed opzioni rispetto alle DTD SGML.

¹⁴ La validazione è il processo mediante il quale si verifica la conformità sintattica della codifica del documento XML con i vincoli imposti nella DTD. Oltre a controllare il corretto utilizzo degli elementi il parser (il software che esegue la validazione) verifica che nell'istanza XML non siano presenti caratteri non supportati dal set dichiarato in apertura di documento.

¹⁵ HTML - HyperText Markup Language - Sito ufficiale W3C <http://www.w3.org/HTML/>.

2.1.2 TEI

Il consorzio internazionale TEI (*Text Encoding Initiative*)¹⁶, istituito nel 2000, si propone come obiettivo principale lo sviluppo di un modello di codifica standard finalizzato alla rappresentazione dell'informazione testuale ed alla gestione dei dati umanistico-letterari in formato elettronico.¹⁷

La TEI è uno standard potente e flessibile utilizzato per la memorizzazione dei testi letterari e finalizzato all'interscambio di documenti tra piattaforme informatiche differenti. Oltre a definire gli schemi di codifica XML, la TEI, rappresenta un vero e proprio vocabolario¹⁸ stabilito attraverso l'associazione di una DTD. Come riferimento alle modalità d'uso degli elementi definiti nella DTD, sono disponibili delle linee guida nominate Guidelines, reperibili liberamente sul sito del consorzio TEI¹⁹. Queste indicazioni hanno trovato largo consenso nella comunità accademica e sono ormai considerate un punto di riferimento importante, come testimoniano i numerosi progetti²⁰ che ricorrono a questo standard.

Gli elementi offerti dalle Guidelines sono organizzati in una struttura modulare permettendo di effettuare scelte che corrispondono alle proprie esigenze e realizzare schemi di codifica appropriati. Ogni modulo è costituito da elementi, detti *tagset*, organizzati in classi: la classe strutturale racchiude gli elementi relativi alla gerarchia del documento mentre, la classe semantica, include gli elementi descrittivi del testo.

I principali moduli sono:

- *tei*, definisce i componenti che verranno utilizzati per tutti i moduli ovvero, le classi di elementi, le macro e i datatype.

¹⁶ Sito di riferimento: <http://www.tei-c.org>.

¹⁷ Il progetto TEI ha preso l'avvio nel 1988 ed è promosso, organizzato e sostenuto dalle tre più importanti associazioni nel campo dell'informatica umanistica e della linguistica computazionale: l'Association for Computers and the Humanities (ACH), l'Association for Computational Linguistics (ACL), e l'Association for Literary and Linguistic Computing (ALLC).

¹⁸ Un vocabolario è un insieme di tag definiti per un determinato scopo.

¹⁹ TEI Guidelines: <http://www.tei-c.org/Guidelines/>

²⁰ Si citano per ragioni di spazio solo alcuni dei numerosi progetti: Vercelli Book Digitale (2003), progetto coordinato da Roberto Rosselli Del Turco e Codice Pelavicino Digitale (2014) su iniziativa di Enrica Salvatori, con lo scopo di fornire un'edizione diplomatica, interpretativa o critica dei testi, immagini digitali ad alta risoluzione, strumenti di consultazione e analisi linguistica.

- *header*, corrisponde all'intestazione contenente i metadati relativi al documento TEI XML²¹.
- *textstructure*, comprende gli elementi strutturali per qualsiasi tipo di documento²².
- *core*, infine, ingloba elementi per qualsiasi tipo di documento²³.

A seguito del rilascio della versione TEI P5, risalente al 1° novembre 2007, sono stati aggiunti altri moduli: il modulo *msdescription*²⁴, contenente i metadati per la descrizione dei manoscritti, e il modulo *gaiji*²⁵, il quale consente di inserire caratteri e glifi che non rientrano nello standard Unicode.²⁶

La TEI offre, inoltre, elementi per la codifica dell'apparato critico, contenuti nel modulo *textcrit*²⁷, ed elementi necessari alla trascrizione delle fonti primarie, appartenenti al modulo *transcr*²⁸. Quest'ultimo offre la possibilità all'utente di collegare il testo ad un'immagine, attraverso la corrispondenza tra l'attributo *@fac*s e gli elementi *<facsimile>*, *<surface>* e *<zone>*, rappresentanti aree di scrittura delimitate da coordinate²⁹.

Sono di notevole importanza anche gli elementi adoperati per la creazione di un'edizione diplomatica, utili per codificare gli interventi editoriali. *<add>* marca le parti di testo aggiunte in origine da un autore, uno scriba o un curatore. *<gap/>* è un elemento milestone, che indica la presenza di una lacuna. *<supplied>* corrisponde ad inserimenti da parte di trascrittori o curatori, adottati allo scopo di integrare o facilitare la lettura del testo originale.

²¹ TEI Guidelines, 2 Header: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>

²² TEI Guidelines, 4 Default Text Structure: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html/>.

²³ TEI Guidelines, 3 Elements Available in All TEI Documents: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html/>.

²⁴ TEI Guidelines, 10 Manuscript Description: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>

²⁵ TEI Guidelines, 5 Characters, Glyphs and Writing Modes: <https://www.tei-c.org/release/doc/tei-p5-doc/it/html/WD.html>

²⁶ L'Unicode Consortium è un consorzio senza fini di lucro, istituito per sviluppare, estendere e promuovere l'uso dello standard Unicode. I membri del consorzio rappresentano un'ampia gamma di aziende e di organizzazioni attive nell'information technology, <http://www.unicode.org>.

²⁷ TEI Guidelines, 12, Critical Apparatus: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>

²⁸ TEI Guidelines, 11 Representation of Primary Sources: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>

²⁹ Per maggiori informazioni, consultare la pagina <https://tei-c.org/release/doc/tei-p5-doc/ko/html/PH.html#PHFAX>

<abbr> marca la presenza di un'abbreviazione, brevigrafo. Al suo interno possono essere presenti elementi <am> utili per specificare le lettere o i segni di abbreviazione. Infine, <expan> contiene un'espansione di un'abbreviazione. Può anche includere gli elementi <ex>, i quali marcano le lettere aggiunte in seguito allo scioglimento dell'abbreviazione.

La versione TEI P5³⁰ introduce un nuovo elemento, <choice>, che consente di raggruppare le coppie di elementi appartenenti allo stesso intervento editoriale.

Ad esempio, tramite l'utilizzo di <choice>, e l'aggiunta dei suoi figli <abbr> e <expan>, è possibile riprodurre due versioni alternative dello stesso testo. Una riporterà l'abbreviazione originale, mentre l'altra, fornirà le espansioni inserite dal curatore.

Questo esempio di applicazione degli elementi editoriali rispecchia la particolare funzionalità della TEI di far coesistere in uno stesso documento livelli di analisi e interpretazioni multiple. Il documento così codificato potrà contenere in contemporanea livelli di edizioni differenti (diplomatica, interpretativa e critica). Ciò è espresso, esplicitamente dall'autore nelle note al testo o internamente all'apparato critico, se presente.

2.2 Edizioni digitali

Per edizione digitale si intende un'edizione di un testo (spesso risalente al passato) preparata e pubblicata seguendo principi e metodi rigorosi. All'interno di esso si implica la presenza di giustificazioni e argomenti proposti dall'editore, a supporto delle sue scelte (edizione critica). Il ruolo del digitale consiste nell'usufruire della sua funzione trasformativa sia nella produzione delle edizioni, sia per quanto riguarda il loro utilizzo. L'edizione digitale può essere vista come un nuovo prodotto che ha come obiettivo quello di permettere la diffusione di contenuti in un formato elettronico. Grazie ad essa, l'utente finale avrà a disposizione particolari strumenti per condurre analisi più approfondite ed elaborate richiedendo una quantità di tempo minima.

³⁰ TEI P5 Guidelines <http://www.tei-c.org/Guidelines/P5/>.

2.2.1 Tipologie

Esistono tre differenti tipi di edizione: il facsimile digitale, l'edizione ipertestuale e l'edizione image based.

L'edizione ipertestuale è stata una delle prime ad essere stata sperimentata nel contesto digitale. Prevede una rappresentazione, generalmente in formato HTML, dei fenomeni codificati in un documento XML TEI, legandoli tra loro attraverso una navigazione ipertestuale. In seguito alla possibilità di avere a disposizione dei repository di immagini scalabili, sono stati introdotti i facsimile digitali, ovvero riproduzioni del manoscritto basate solamente su scansioni digitali. Infine, l'edizione image-based, essendo la più completa fino ad oggi, associa alla trascrizione del testo codificato delle riproduzioni facsimilari composte da immagini del manoscritto. Oltre alla classificazione appena presentata, le edizioni digitali, possono essere suddivise in tre ulteriori categorie: l'edizione diplomatica, l'edizione interpretativa e l'edizione critica, le quali possono coesistere all'interno dello stesso documento. L'edizione diplomatica riproduce il testo allo stesso modo in cui appare nel manoscritto, comprendendo eventuali errori, abbreviazioni, lettere diverse o non standard.

L'edizione interpretativa modernizza il contenuto testuale, adattandolo ad un utilizzo specifico e interpretandone i segni grafici (unisce o separa le parole, scioglie le abbreviazioni, aggiunge apostrofi e accenti, rivede la paragrafatura e segnala gli errori e le lezioni sospette). Infine, l'edizione critica ricostruisce il testo in modo che sia quanto più vicino possibile all'originale. Essa per essere completa deve offrire un apparato critico, un'introduzione paleografica/filologica, dei commenti esplicativi e una bibliografia di riferimento.

Un'edizione digitale è composta quindi come minimo dalla trascrizione testuale e da un software di visualizzazione al quale si possono aggiungere scansioni ad alta risoluzione del testo originale.

3 EVT-Edition Visualization Technology

Intorno al 2008, all'interno del corso di Informatica Umanistica dell'Università di Pisa, si ebbe la necessità di trovare una soluzione specifica avente come obiettivo l'elaborazione di un software per la visualizzazione di edizioni digitali. Il progetto fu guidato dal professore Roberto Rosselli Del Turco e venne intrapreso da un buon numero di studenti. Le prime due versioni di EVT presentavano problemi di vario tipo, uno su tutti coincideva con l'inserimento manuale dei dati all'interno dell'interfaccia di presentazione, nel quale dovevano essere specificati i nomi di tutti i file che componevano l'edizione. Questo svantaggio richiedeva non solo costi in termini di fatica da parte del programmatore, ma anche, numerosi aggiornamenti dovuti alle variazioni del file. La svolta logica per ovviare a questa difficoltà si ebbe rovesciando il punto di vista adottato fino a quel momento: non sono i file che devono essere inseriti nell'applicazione, ma è l'applicazione stessa a dover essere generata a partire dai file dell'edizione. Nella prima versione di EVT, infatti, il punto di partenza è il documento TEI XML che contiene i dati dell'edizione. Ad esso viene applicato un foglio di stile XSLT³¹, grazie al quale è possibile avviare una catena di trasformazioni mutando il nostro file in una applicazione Web pronta per essere caricata su un server.

Nato come soluzione di uno specifico progetto, EVT oggi è un software flessibile per la creazione e consultazione di edizioni digitali (manoscritti antichi e medievali) che si propone come obiettivo principale quello di essere uno strumento che consente la navigazione e lo studio di testi, attraverso un'interfaccia intuitiva e user-friendly. Di seguito riportiamo le caratteristiche principali possedute da EVT:

- Il software è multiplatforma, in grado quindi di essere eseguito dai sistemi operativi più comuni.
- È distribuito con licenza Open Source, essendo quindi accessibile a tutti.
- È espandibile, permettendo di elaborare future versioni aggiornate.

³¹ XSL Transformations (XSLT) è un linguaggio dichiarativo utilizzato per manipolare documenti XML. Esso estrae i dati da un documento XML per produrre un nuovo documento. Il processo di creazione di un nuovo documento a partire da un sorgente XML è detto *trasformazione*, questo processo non cambia il documento originale. Cfr. XSL Transformations - Sito ufficiale W3C - <https://www.w3.org/TR/xslt/>.

- Possiede un'interfaccia intuitiva, integrata nel sistema operativo su cui viene eseguita.
- EVT ha la capacità di gestire XML, visualizzati con fogli di stile XSLT³², e file HTML.
- Permette l'integrazione di un motore di ricerca XML.

Grazie a queste sue proprietà, il software EVT, ha avuto un grande successo nel campo dell'editoria digitale venendo utilizzato all'interno di vari progetti.

3.1 Funzionalità

EVT, a partire già dalla versione 1.0, mette a disposizione le seguenti funzionalità principali.

- Visualizzazione simultanea dei testi e delle immagini in due riquadri separati, dando la possibilità di scegliere all'utente la modalità di visualizzazione (testo-testo, testo-immagine o solo immagine).
- Possibilità di consultare, anche in contemporanea, più livelli di edizione del testo: diplomatica, diplomatica-interpretativa e critica.
- Presenza di un collegamento testo-immagine.
- Utilizzo dello zoom.
- Presenza di un motore di ricerca XML, che consente la gestione e l'identificazione delle named entities³³ come ad esempio i nomi di persona, le date, i luoghi etc...
- Infine, il funzionamento del software è totalmente in locale (client side) e permette la condivisione e la diffusione dell'edizione senza la necessità di avere un server di appoggio.

Oltre alle funzioni appena descritte, è giusto menzionare la presenza di una documentazione elettronica del software, situata sulla pagina github, utile per configurare ed adoperare EVT.

³² XSLT è formalmente un vocabolario XML, si presenta infatti come un insieme di tag XML, i quali non dicono ad un programma come visualizzare qualcosa, ma piuttosto cosa fare quando incontra un certo tag.

³³ La gestione in EVT delle named entities è stata sviluppata per il Codice Pelavicino; questa funzionalità consente di creare delle liste dei nomi di persona, dei luoghi e delle organizzazioni presenti nel testo.

3.2 Seconda Versione

Negli anni successivi allo sviluppo della prima versione del software EVT 1, si era giunti alla consapevolezza di dover integrare delle nuove caratteristiche allo scopo di supportare edizioni critiche complesse. Il codice XSLT 2 che costituiva il nucleo centrale di EVT non era solo cresciuto considerevolmente, ma era divenuto complicato e piuttosto difficile da modificare. Anche il numero degli elementi grafici era ugualmente aumentato, rendendo il layout dell'applicazione altamente confusionario. È per tutti questi motivi che si decise di evitare l'utilizzo dei fogli di stile XSLT, rendendo così il software più leggero, flessibile e modulare. Si scelse quindi di basare il nuovo framework di sviluppo sull'architettura Model View Controller (MVC), un design pattern adottato nella programmazione ad oggetti ed implementato dal framework Angular JS.

I principi di funzionamento restano invariati, dato che anche EVT 2 accetta come input edizioni codificate usando lo standard XML TEI, ma la struttura del software è stata progettata in maniera tale da poter integrare anche un parser per edizioni preparate con altri linguaggi di markup. Per quanto riguarda la flessibilità e modularità del software, basterebbe questa caratteristica a rendere EVT 2 migliore rispetto alla versione precedente, condizionata dai limiti di XSLT.

3.3 Architettura

Il software EVT si presenta come un set completo di strumenti, flessibili e altamente personalizzabili, sviluppato per consentire agli utenti di visualizzare, leggere e confrontare le edizioni in un ambiente elettronico.

Si basa su tecnologie web standard aperte, come XML, HTML³⁴, CSS³⁵ e Javascript³⁶ che garantiscono il funzionamento su tutti i browser web più recenti e la durabilità sul World Wide Web.

L'architettura del software è stata rimodellata, in modo tale da renderlo più leggero, usabile e adattabile, usufruendo dell MCV che, come abbiamo già espresso nel paragrafo precedente, è un pattern architetturale molto comune nella progettazione ad oggetti. Esso permette di separare il processo di presentazione dei dati dalla logica applicativa e dal nucleo di elaborazione, sfruttando il framework JavaScript AngularJS che consente a sua volta di definire componenti HTML personalizzati e di associare il modello degli elementi dell'interfaccia utente.

Prima di questa rielaborazione della struttura, EVT era composto da due unità principali: EVT Builder per la trasformazione del testo codificato, tramite appositi template XSLT 2.0, e EVT Viewer per la visualizzazione sul browser dei risultati della trasformazione e l'interazione con essi. La nuova versione di EVT ha quindi il compito di leggere, parsare il file del testo codificato in XML e salvare gli elementi all'interno di un modello dati che rimane client side nella memoria principale. In seguito a queste modifiche, si è deciso quindi di eliminare EVT Builder così da poter permettere allo studioso di aprire la propria edizione digitale direttamente sul browser, senza dover usufruire delle trasformazioni attuate da Saxon.

³⁴ L'HTML (HyperText Markup Language) è il linguaggio di marcatura per ipertesti# che consente di formattare e collegare tra loro le pagine web.

³⁵ I fogli di stile CSS (Cascading Style Sheets) definiscono la modalità di visualizzazione degli elementi che costituiscono una pagina.. Sito ufficiale W3C: <https://www.w3.org/TR/css/>.

³⁶ JavaScript è un linguaggio di programmazione utilizzato dagli sviluppatori per creare pagine Web interattive. Sito ufficiale JavaScript <https://www.javascript.com/>.

4 Progetto di codifica della Pace di Dante

Il progetto di codifica della Pace di Dante ha avuto diverse fasi di sviluppo non sempre sequenziali. Una prima fase è stata di tipo concettuale, si è concordato un itinerario del progetto stabilendo le modalità di codifica da adottare e le procedure da seguire. Dopodiché si è svolto un lavoro di controllo della trascrizione dei manoscritti dalle immagini fornite direttamente dall'Archivio di Stato della Spezia. La fase successiva ha riguardato il processo di marcatura del testo e, più in generale, l'elaborazione del documento XML TEI. Infine, si è impostata la visualizzazione del documento digitale in accordo alle rigide regolamentazioni della sintassi di EVT.

4.1 Pianificazione concettuale

L'inizio vero e proprio del progetto è stato preceduto da una fase preparatoria durante il quale sono stati definiti gli obiettivi, le risorse necessarie per avviarlo e, soprattutto, sono stati investigati i vari strumenti richiesti per le diverse operazioni da portare a termine. Si è optato quindi per l'elaborazione di un'edizione diplomatica-interpretativa delle pagine appartenenti alla Pace di Dante, fornita di testo codificato e immagini corrispondenti del manoscritto originale. L'edizione critica delle medesime pagine, sviluppata durante il progetto di tirocinio, ha funto da modello di riferimento per le successive fasi del progetto. Da essa, infatti, sono stati ripresi e, in seguito, riadattati i metadati relativi al documento XML TEI e, inoltre, si è decretato il modello di visualizzazione che avrebbe dovuto assumere il documento. A fronte del layout confusionario, dovuto al gran numero degli elementi grafici elaborati dai fogli di stile XSLT 2, si è deciso di prediligere l'utilizzo del software di presentazione EVT 2. Questo ha comportato uno studio approfondito dell'applicazione, dalla sua configurazione più generale, sino alla comprensione dettagliata di alcune sue specifiche.

4.2 Digitalizzazione delle immagini e trascrizione dei testi

Come già accennato, l'Archivio di Stato della Spezia ha messo a disposizione dodici immagini ritraenti le *tabulae* della Pace di Dante, suddivise a loro volta in recto e verso. Il progetto ha riguardato, in particolare, la realizzazione di un file digitale dalla pagina 270 recto alla pagina 275 verso. I formati digitali delle pagine sono in formato JPEG e la loro risoluzione (7057x4724) è stata adattata, utilizzando il software Adobe Photoshop³⁷, allo scopo di ottenere una loro uniforme visualizzazione finale in EVT 2. Dopo aver modificato le immagini, si è proseguito con la trascrizione del testo utilizzando il programma di video-scrittura Word. In esso, il contenuto testuale è stato suddiviso in pagine e a sua volta in linee, facendo riferimento all'edizione critica realizzata durante il progetto di tirocinio. Quest'ultima è stata ottenuta dalla digitalizzazione dei documenti, dal 145 al 148, appartenenti all'edizione del Codice Diplomatico Dantesco del 2016. Dall'edizione critica sono stati ripresi gli interventi editoriali e le note ai testi adottate dai curatori della raccolta: Teresa De Robertis, Giuliano Miliani, Laura Regnicola e Stefano Zamponi.

4.3 Criteri di edizione e markup

Terminata la trascrizione dei manoscritti, si è stabilito quali fossero le caratteristiche del testo da marcare per definire il modello di codifica da adottare. La TEI offre schemi di codifica di vario tipo per testi in ambito umanistico, lasciando al codificatore la libertà di poter attuare una personalizzazione tramite l'inserimento di moduli ed elementi opzionali³⁸. Una volta selezionato lo schema di codifica "tei all", contenente tutti i moduli base, si è verificata la sua validità e, infine, si è controllato di avere a disposizione tutti gli elementi necessari alla marcatura del testo.

La codifica del testo è stata realizzata in maniera non lineare. Dopo aver valutato quali elementi potevano essere impiegati per la trascrizione dei testi, ci si è dedicati a risolvere

³⁷ Adobe Photoshop è un software proprietario prodotto da Adobe Systems Incorporated specializzato nell'elaborazione di foto e di immagini digitali. <http://www.adobe.com/it/products/photoshop.html>.

³⁸ "un apporto assolutamente soggettivo dello studioso-codificatore, che necessita di interpretare la funzione delle varie componenti strutturali tipografiche, o manoscritte, per essere in grado di impiegare in modo adeguato i marcatori per gli elementi previsti nella DTD della TEI" Cfr. Fabio Ciotti, *Breve introduzione alla Text Encoding Initiative*, in *Biblioteca Italiana* http://www.bibliotecaitaliana.it/tei_intro.asp, 1 dicembre 2003.

eventuali problemi specifici dovuti a particolarità del manoscritto, man mano che essi si presentavano.

Il processo di codifica vero e proprio è iniziato con la creazione del file XML contenente il testo e la marcatura, adattata allo schema di codifica TEI precedentemente definito.

Le pagine, contenenti i cinque documenti, sono state raggruppate all'interno del corpo del testo, contrassegnando il loro inizio tramite l'elemento milestone `<pb/>`³⁹ e assegnandogli l'attributo `xml:id` per differenziarle tra loro.

Il file, nominato `Dantetesi.xml`, contiene una serie di elementi standard organizzati in una struttura gerarchica ad "albero". La radice è rappresentata dall'elemento `<TEI>`, associato al namespace "<http://www.tei-c.org/ns/1.0>", il quale include i macro elementi che costituiscono la struttura del documento XML. L'elemento `<teiHeader>` corrisponde all'intestazione del documento e contiene i metadati relativi alla descrizione dell'edizione digitale. Al suo interno troviamo l'elemento `<fileDesc>` che comprende a sua volta quattro principali elementi : il primo, `<titleStmt>`, fornisce le informazioni relative al titolo dell'opera (`<title>`), dell'autore (`<author>`) e del responsabile del documento digitale (`<respStmt>`). Il secondo elemento, `<editionStmt>`, racchiude le informazioni relative alla specifica edizione digitale, integrando i responsabili dell'edizione critica, da cui ci si è basati per le integrazioni alle lacune. Il terzo elemento, `<publicationStmt>`, contiene le informazioni sulla pubblicazione dell'edizione digitale. Infine, il quarto elemento, `<sourceDesc>`, descrive le caratteristiche specifiche delle fonti primarie. In esso sono presenti una lista di persone (`<listPerson>`) e di luoghi (`<listPlace>`) menzionati nel testo. La lista `<listPerson>` contiene una serie di elementi `<person>`, ciascuno identificato dall'attributo `@xml:id`, e altri metadati relativi alla persona (sesso `<sex>`, data di nascita e di morte `<birth>` e `<death>` e altre informazioni aggiuntive). La lista `<listPlace>` contiene gli elementi `<place>` identificati anch'essi dall'attributo `@xml:id` e altri metadati per la descrizione delle località. Il modulo TEI P5 `msdescription` consente di indicare le caratteristiche delle fonti documentarie all'interno dell'elemento `<msDesc>` contenuto anch'esso dal `<sourceDesc>`.

³⁹ `<pb/>` è un tag autonomo, vuoto, che non prevede un tag di chiusura. TEI Guidelines: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pb.html>.

Al suo interno troviamo l'elemento `<msIdentifier>` che contiene alcune informazioni utili all'identificazione del manoscritto: il paese di provenienza, la regione, la città, il nome dell'ente che lo conserva, la descrizione della raccolta che lo comprende e il nome. L'elemento `<physDesc>`, invece, contiene la descrizione della componente fisica, indicandone il materiale, la composizione e la condizione attuale del manoscritto.

`<msDesc>` è suddiviso a sua volta da cinque `<msPart>`, ciascuno relativo ad un documento, così da poter implementare informazioni maggiormente dettagliate per ciascun singolo atto. Un'altro componente del `<teiHeader>` è l'`<encodingDesc>` che descrive le modalità adottate per codificare le diverse tipologie degli scioglimenti di abbreviazioni. Allo stesso livello gerarchico del `<teiHeader>` abbiamo l'elemento `<facsimile>`, introdotto dalla TEI negli schemi di codifica della versione P5. Grazie alla presenza del modulo *transcr*⁴⁰ è stato possibile realizzare la parallel transcription, ossia il collegamento tra trascrizione del testo e immagine del manoscritto. All'interno dell'elemento `<facsimile>` troviamo l'elemento `<surface>`, la superficie di scrittura, collegato tramite l'attributo `@fac` all'`xml:id` degli elementi rappresentanti gli inizi delle pagine (`<pb />`). Gli elementi che compongono il `<surface>` sono il `<graphic>` che, grazie alla presenza dell'attributo `@url`, fornisce l'immagine della pagina assegnandone il percorso⁴¹ corrispondente e le numerose `<zone>`, le quali delineano, tramite coordinate cartesiane dell'immagine, le aree di interesse per ciascuna linea del manoscritto.

Per collegare gli elementi `<zone>` al testo corrispondente si utilizza, negli elementi testuali, l'attributo globale `@fac` che contiene il riferimento all' `xml:id` degli elementi `<surface>` per le pagine, e degli elementi `<zone>` per le linee del testo. Infine la struttura centrale del testo del documento è racchiusa dall'elemento `<text>`⁴², il corpo del file, e a sua volta dall'elemento `<group>` che raggruppa il contenuto dei cinque documenti dando l'opportunità di realizzare un documento composito.

⁴⁰ TEI Guidelines, 8 Transcriptions of Speech:

<https://www.tei-c.org/release/doc/tei-p5-doc/it/html/TS.html#HD32>

⁴¹ Il percorso può essere di tipo assoluto, ossia viene descritto tutto il percorso da seguire lungo l'albero del documento a partire dall'elemento radice, oppure relativo facendo riferimento direttamente al nodo contesto dell'elaborazione.

⁴² L'elemento `<text>` può rappresentare un testo unitario o un testo composito. Quest'ultimo corrisponde a una collezione di opere e prevede un ulteriore elemento figlio `<group>` il quale è da impiegarsi nel caso di una serie di testi facenti parte della medesima pubblicazione, ognuno dotato della propria individualità (ciascuno con un proprio front, body e back).

I cinque differenti atti sono stati rielaborati in cinque strutture tra loro univoche. Sono presentati, infatti, cinque elementi `<text>`, ciascuno relativo ad un documento, che a loro volta possiedono gli elementi `<front>` e `<body>`.

L'elemento `<front>`, contenente le diverse voci bibliografiche racchiuse negli elementi `<bibl>`⁴³, fornisce una serie di dettagli che confluiscono nel regesto dello specifico documento. All'interno di `<bibl>` troviamo per esempio la datazione, l'intestazione, il titolo e la bibliografia attinente alle varie edizioni dei documenti. In ciascun regesto sono stati marcati i nomi di persona tramite l'elemento `<persName>` e l'attributo `@ref`, il quale rimanda all'`@xml:id` dell'elemento `<person>` contenuto in `<listPerson>`. Allo stesso modo, i nomi dei luoghi sono stati contrassegnati utilizzando gli elementi `<placeName>` e l'attributo `@ref`, collegato all' `@xml:id` della lista dei luoghi.

Come già accennato, l'elemento `<body>` è posizionato allo stesso livello dell'elemento `<front>` e contiene il vero contenuto testuale dei documenti. All'interno di esso si è inserito l'elemento `<ab>`, il quale funge da contenitore anonimo per porzioni di testo di cui non si vuole esplicitarne il valore semantico. Questa soluzione è stata impiegata con lo scopo di impedire la presenza di gerarchie sovrapposte e, quindi, per rendere il documento valido in accordo allo schema fornito dal consorzio TEI.

Una volta creata la struttura fondamentale per la codifica del testo di ciascun atto sono stati aggiunti alla trascrizione del testo due tipologie di elementi milestone a livello strutturale: l'elemento `<pb/>` rappresenta l'inizio pagina di un documento, mentre, l'elemento `<lb/>`⁴⁴ corrisponde alla suddivisione delle righe. Questi ultimi due elementi presentano due attributi, `@xml:id` e `@n`, che li identificano univocamente rendendo possibile sia il collegamento tra le numerose `<zone>`, che la loro visualizzazione in EVT. Alcuni elementi `<lb/>` contengono anche l'attributo `break` impostato con valore "no", che fornisce indicazioni laddove la sillabazione di una parola viene spezzata dall'andare a capo.

4.4 Livelli di edizione del testo

⁴³ Tali elementi contengono una citazione bibliografica non strutturata, i cui componenti possono, o meno, essere marcati esplicitamente.

⁴⁴ L'elemento `<lb/>` (line beginning) è un elemento milestone, che indica l'inizio di una nuova riga tipografica. TEI Guidelines: <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-lb.html>

4.4.1 Edizione diplomatica-interpretativa

La prima fase di sviluppo del progetto ha previsto l'elaborazione dell'edizione diplomatica-interpretativa, con l'obiettivo di riprodurre il testo allo stesso modo in cui appare nel manoscritto, realizzare il collegamento testo-immagine e renderlo fruibile per la visualizzazione in EVT 2. Non avendo un'altra edizione tradizionale del manoscritto da cui far riferimento, si è deciso di realizzarne una nuova identificando i documenti e dotandogli di un identificativo numerico, dal numero uno al numero cinque. Il markup integrato riprende le convenzioni dell'edizione critica svolta durante il lavoro di tirocinio inglobandone di nuove, essenziali alla realizzazione dell'edizione diplomatica-interpretativa. Quest'ultima, infatti, possiede una maggiore quantità di interventi editoriali: viene rispettata la divisione in righe, in pagine e quella dei documenti; vengono indicate le abbreviazioni e le espansioni; vengono marcate le lacune e, infine, vengono mostrati i glifi ritenuti non convenzionali al nostro linguaggio.

Per la ripartizione dei documenti, in accordo con le rigide regole del software EVT, si è deciso di utilizzare cinque elementi `<text>`, i quali contengono l'attributo identificativo univoco `@xml:id` con valore associato al numero del documento (dal *doc_1* al *doc_5*).

```
<text xml:id="doc_1">
```

La suddivisione delle pagine è stata sviluppata inserendo 13 elementi milestone `<pb />`, ciascuno corrispondente ad una pagina del manoscritto.

```
<pb n="1RB" xml:id="Dante_fol_1RB"  
fac="data/images/single/Tabula1_RB.jpg"/>
```

All'attributo `@n` è stato associato un nominativo per differenziare tra loro le pagine del manoscritto. Esse sono ripartite in base al recto o verso ("R" o "V") e all'impaginazione (A destra e B sinistra).

L'identificazione univoca dell'elemento `<pb/>` è stata possibile grazie all'utilizzo dell'attributo `@xml:id`, mentre, l'attributo `@fac`, è stato adoperato per la creazione del collegamento al percorso dell'immagine del manoscritto.

Durante lo sviluppo di questa tipologia di struttura è sorto il problema derivato dalla sovrapposizione di gerarchie. In special modo, si è manifestata la necessità di dover suddividere la pagina *6RB*, contenente la fine del *doc_3* e l'inizio del *doc_4*, sia per documento che per pagina. Per ovviare a questo problema sono state adoperate delle modifiche al file xml riguardanti la duplicazione della pagina *6RB* e la rinominazione delle due parti in *6RBF* e *6RBI*. La prima rappresenta la parte finale del documento 3, mentre, la seconda, corrisponde all'inizio del documento 4. Nella visualizzazione questa loro particolarità è messa in risalto tramite una variazione cromatica⁴⁵ volta a differenziare il contenuto testuale dei due documenti.

L'ultimo elemento strutturale per il corpo centrale del documento è il milestone `<lb/>`, utilizzato per suddividere il testo in righe in maniera speculare rispetto al manoscritto originale.

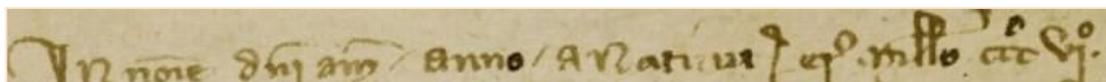


Figura 1: Prima riga della pagina 1_RB

```
<lb n="1" xml:id="Dante_lb_1RB_01" facs="#Dante_line_1RB_01"/>
```

A questo elemento sono associati l'attributo `@n`, per il numero di riga, l'identificativo `@xml:id`. L'attributo `@facs`, invece, è stato sfruttato per costruire il collegamento con l'`xml:id` delle zone, le quali rappresentano aree dell'immagine contenenti le righe del manoscritto. Alcuni elementi `<lb/>` presentano anche l'attributo `break`, utile per marcare lo spezzarsi di una parola cusato dall'andare a capo.

Per quanto riguarda le abbreviazioni, le espansioni e la marcatura di caratteri aventi una propria simbologia semantica, si sono sfruttate tre diverse metodologie:

- La maggior parte delle abbreviazioni sono state marcate dall'elemento `<expan>`. In particolare, si sono trattate, con questa modalità di codifica, le parole che presentano

⁴⁵ Nel file CSS sono specificati i valori in inglese dei colori tramite il modello RGBA.

una mancanza di alcuni caratteri rispetto alla loro originaria espansione. All'interno di `<expan>` le lettere omesse sono designate dagli elementi `<ex>`.

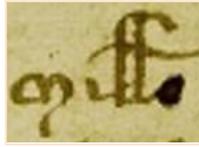


Figura 2: Parola abbreviata con mancanza di caratteri

```
<expan>mill<ex>esim</ex>o</expan>
```

- Le abbreviazioni che contengono simboli corrispondenti a lettere, ma con evidente funzione di abbreviazione, sono state marcate con l'elemento `<choice>`. All'interno di quest'ultimo avremo la coppia degli elementi `<abbr>` e `<expan>`, i quali identificano rispettivamente la parola abbreviata e la sua corrispettiva espansione. L'elemento `<am>`, interno ad `<abbr>`, servirà a codificare il carattere che simboleggia l'abbreviazione, mentre, l'elemento `<ex>` esalterà le lettere omesse.

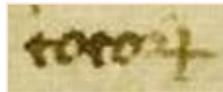


Figura 3: Parola abbreviata con simbolo che sostituisce lettere

```
<choice>  
  <abbr>totor<am>+</am></abbr>  
  <expan>totor<ex>um</ex></expan>  
</choice>
```

- Infine, le abbreviazioni che contengono glifi, con chiara funzione di rimpiazzo in insiemi di lettere, sono codificate con l'elemento `<choice>` e i suoi due figli: `<reg>` e `<orig>`. L'elemento `<orig>` serve per identificare la forma originaria della parola, al contrario, l'elemento `<reg>` viene utilizzato per identificare una lettura che è stata regolarizzata o normalizzata.



Figura 4: Parola abbreviata contenente un glifo

```
<choice>
  <orig>g</orig>
  <reg>contra</reg>
</choice>
```

Un'altro aspetto importante inerente alla codifica dell'edizione diplomatica-interpretativa è stato l'inserimento degli interventi editoriali. Per la risoluzione di questa tematica ci si è basati sul lavoro svolto da Teresa De Robertis, Giuliano Milani, Laura Regnicoli e Stefano Zamponi, i curatori dell'edizione del 2016 del codice diplomatico Dantesco.

Inizialmente ci si è dedicati alla codifica di due tipologie di integrazioni: quelle utilizzate per porre rimedio all'impossibilità di lettura causata dal guasto al supporto materiale e, quelle volte ad ovviare l'omissione di parti di testo del manoscritto originale. Tramite l'elemento `<supplied>`, e l'associazione di alcuni suoi attributi, siamo stati in grado di marcare le integrazioni, differenziandole tra loro e esplicitando i responsabili grazie all'utilizzo dell'attributo `@resp`. Quest'ultimo è collegato all'identificativo della persona, contenuta nella lista `<listPerson>`, responsabile dell'integrazione.

- Per le integrazioni su guasto del supporto è stato associato all'attributo `@reason`, appartenente a `<supplied>`, il valore "support_damage".

```
<supplied resp="#TD" reason="support-damage">
  integrazioni per guasto del supporto
</supplied/>
```

- Per le integrazioni di lettere o parole omesse per errore nel testo, invece, all'attributo @reason è stato assegnato il valore "omitted-in-original"

```
<supplied resp="#LR" reason="omitted-in-original">
    Integrazione di lettere o parole omesse nel testo
</supplied>
```

Allo scopo di ampliare il bagaglio editoriale dell'edizione diplomatica-interpretativa, si è deciso di inserire gli elementi <gap>, i quali coincidono con le lacune del testo scaturite da guasti al supporto. L'attributo @style è stato assegnato per definire il posizionamento della lacuna: a destra, a sinistra o al centro. Normalmente, in accordo con le linee guida della TEI, l'elemento <gap> è riconosciuto come elemento milestone, e quindi privo di contenuto. Nel nostro documento digitale, però, si è deciso di assegnargli l'elemento figlio <desc>, adeguando la sintassi alle rigide regole di visualizzazione del software EVT.

```
<gap style="right" reason="illegible" agent="damage">
    <desc>&#x2026;</desc>
</gap>
```

Infine sono state aggiunte le note ai testi, sempre facendo riferimento all'edizione del 2016 del Codice diplomatico dantesco. Le note, richiamate da lettere, giustificano la lezione a testo o rendono conto dello stato dell'originale quanto a correzioni, omissioni, lacune, incertezze di lettura, ecc.; riportano le lezioni significative di altri eventuali testimoni e, in casi eccezionali, le letture differenti di altri editori. In quest'ultimo caso, la variazione testuale viene seguita da una parentesi quadra chiusa e in seguito vengono elencate le diverse edizioni che concordano con essa.

```
<note type="comment" n="a">
    dictum dominum] dominum dominum Maccioni, Vercelli,
    Vernon, Fraticelli
</note>
```

4.4.2 Edizione critica

Nell'inserimento di elementi di codifica per l'edizione diplomatica e interpretativa ci si è basati sul lavoro svolto in precedenza durante il progetto di tirocinio. Quest'ultimo ha riguardato l'elaborazione di un'edizione critica dei documenti della Pace di Dante contenuti nel libro "Nuova Edizione commentata delle Opere di Dante" del 2016. In un primo momento, sono stati presi in esame gli interventi editoriali curati da Teresa De Robertis, Giuliano Milani, Laura Regnicola e Stefano Zamponi e, successivamente, ci si è dedicati all'elaborazione di un nuovo documento digitale. L'organizzazione strutturale dei cinque documenti (dal numero 144 al 148) è stata ottenuta dalla creazione di un documento composito formato da cinque elementi `<text>`, ciascuno contenente un elemento `<front>` (per il regesto), un `<body>` (per il contenuto testuale vero e proprio) e infine un elemento `<back>`. Il corpo del testo è stato suddiviso in paragrafi mediante l'utilizzo degli elementi `<p>`, rispecchiando la composizione dell'edizione del 2016.

Per la visualizzazione di questa tipologia di struttura, sono stati inseriti degli elementi milestone, i quali sfruttano il valore dell'attributo `@n` per dare identificazione del numero di paragrafo.

Oltre agli interventi editoriali, già menzionati ed adottati per la realizzazione dell'edizione diplomatica-interpretativa, va sottolineata la marcatura dei nomi di persona e dei luoghi, realizzata tramite l'inserimento di `<persName>` e `<placeName>`. Questi due elementi, a differenza del progetto finale, sono presenti non solo all'interno dei regesti, ma sono anche contenuti dai vari paragrafi del contenuto testuale. La differenza sta nel fatto che l'edizione del 2016 si mostra nella sua forma più completa, essendo munita dello scioglimento delle abbreviazioni, di illustrazioni tramite note ai testi e delle integrazioni da parte dei suoi curatori. Per questo motivo i nomi delle persone e dei luoghi possono essere identificati con maggior chiarezza rispetto alla nostra edizione diplomatica-interpretativa tratta dal manoscritto originale, senza aver la possibilità di cadere in problemi di ambiguità.

L'aspetto peculiare dell'edizione critica è sicuramente la composizione di un'apparato critico situato all'interno dei differenti `<back>`, ciascuno associato ad un singolo documento. Per raggruppare le differenti voci di un documento, marcate dall'elemento

<app>, sono state adoperate le liste <listApp>, le quali possiedono l'attributo @type, utile ad identificare l'associazione a uno specifico documento.

Gli elementi <app> sono collegati tramite due attributi di posizione (@from e @to) ad un elemento ancora (<anchor>) interno al corpo del testo. Le varie voci dell'apparato critico sono identificate da due tipologie di elementi contenuti da <app>: l'elemento <lem>, che racchiude il lemma o testo base, e l'elemento <rdg>, il quale contiene una singola lettura coincidente a una variazione testuale.

In alcuni casi si è dovuto soddisfare la necessità di raggruppare le singole letture appartenenti alla stessa variazione testuale. Si è dunque scelto di inserire le diverse versioni testuali all'interno dell'elemento <rdgGrp>, e di collegarle, tramite l'attributo @source, al corrispondente riferimento bibliografico contenuto in <listbibl>.

La visualizzazione dell'edizione critica ha richiesto delle elaborazioni maggiormente flessibili rispetto a quelle adottate nell'edizione diplomatica-interpretativa, dove si è dovuto rispettare le rigide regolamentazioni del software EVT 2. La parte finale del progetto ha riguardato la creazione di un file XSLT ricco di regole di trasformazione, le quali ci hanno permesso di convogliare gli elementi, precedentemente marcati nel file XML, in un nuovo documento digitale presentato tramite l'HTML. Per quanto riguarda l'aspetto finale assunto dall'edizione, sono stati inseriti un file CSS e un file Javascript, così da poter rendere la pagina maggiormente interattiva e migliorarne la resa grafica.

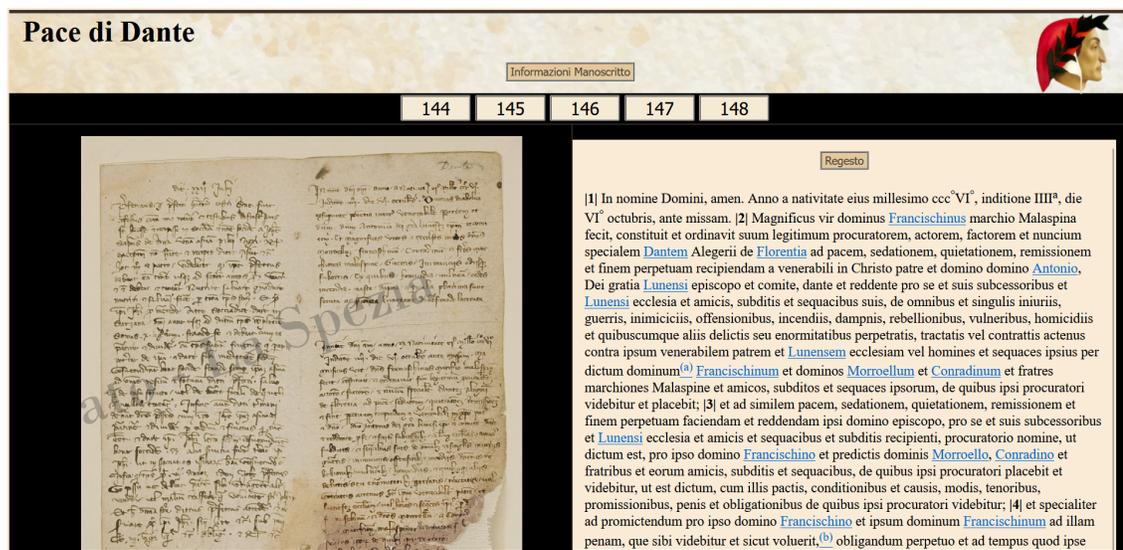


Figura 5: Home page dell'edizione critica Pace di Dante

Nella sua rappresentazione finale, il file digitale si presta ad una visualizzazione di tipo testo-immagine implementata dalla suddivisione dei cinque documenti. In particolare a sinistra, viene raffigurata l'immagine del manoscritto originale, mentre, a destra è possibile osservare il testo trascritto coincidente con l'edizione del 2016 del Codice diplomatico Dantesco. Per non incorrere in problemi di ambiguità, si è deciso di non adoperare il collegamento testo-immagine siccome il testo proveniente dall'immagine del manoscritto originale presenta incongruenze rispetto alla trascrizione del testo dell'edizione del 2016.

Per quanto riguarda l'aspetto critico, sempre tramite le regole trasformazione del file XSLT, è stato possibile mettere in evidenza le varie voci di apparato, collegandole tramite ancoraggio alla loro posizione interna al testo, identificata da una lettera.

In una sezione sottostante alla trascrizione del testo sono presenti degli elementi button che permettono all'utente di interagire e mettere in risalto i vari elementi editoriali marcati all'interno del documento XML. In specifico, possiamo evidenziare: le parole omesse nel testo del manoscritto, le integrazioni adottate a fronte del guasto al supporto, le espansioni e il loro scioglimento e, infine, i nomi di persone e di luoghi presenti nel testo.

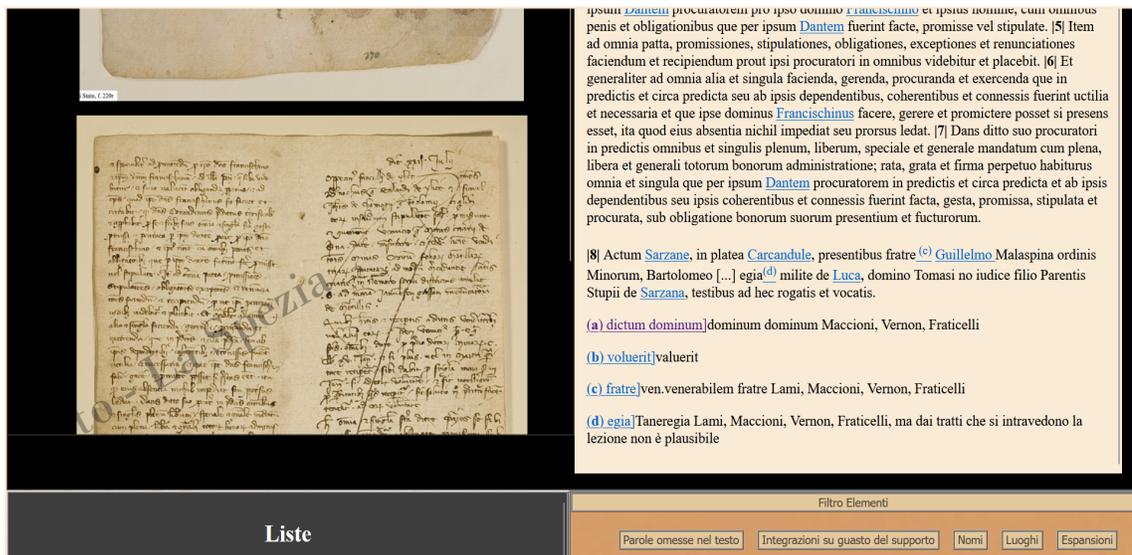


Figura 6: Apparato critico e selettori di named entities

4.5 Dal processo di codifica alla visualizzazione

La fase finale del progetto ha riguardato l'elaborazione della visualizzazione dell'edizione diplomatica-interpretativa corrispondente ai documenti della Pace di Dante. Essa è stata pensata diversamente rispetto alla visualizzazione dell'edizione critica, nel quale si è mostrato un file HTML, generato e compilato dalle regole di trasformazioni XSLT (eXtensible Stylesheet Language Transformations). Prendendo quindi in esame il progetto di tirocinio, svolto in precedenza, si è giunti alla consapevolezza di dover optare per una soluzione universale, così da poter rendere l'edizione facilmente implementabile dall'aggiunta di progetti futuri.

Per quanto riguarda la visualizzazione del documento digitale, si è deciso di utilizzare la seconda versione del software EVT, scaricabile alla pagina di riferimento (nota), la quale ci ha permesso di convogliare il risultato finale dell'edizione, tramite un'interfaccia di presentazione, privandoci dell'utilizzo dei fogli di stile XSLT. A fronte della rigidità delle regole che vincolano il funzionamento di EVT 2, si sono dovuti adattare alcuni elementi del file XML, che non venivano riconosciuti dal software. In particolare, gli elementi milestone `<gap>`, rappresentanti le lacune del testo, non erano visualizzati dal software EVT a causa della loro mancanza di contenuto. I `<gap>` sono stati quindi trasformati in elementi con contenuto, associandogli i figli `<desc>` e inserendo al loro interno il carattere Unicode⁴⁶ `'…'`, rappresentante tre punti volti a simboleggiare la mancanza di testo.

In alcuni casi, invece, si è dovuto intervenire attraverso alcune modifiche dei file contenuti all'interno delle cartelle di EVT 2. Prendendo in esame l'organizzazione strutturale in cartelle del software EVT 2, delineiamo adesso, i cambiamenti adottati allo scopo di inserire tutti gli elementi marcati del file XML.

La maggior parte degli interventi ha riguardato il file `'config.json'`, contenente la configurazione del nostro software di visualizzazione. All'interno di esso è stato impostato: il collegamento al file XML, il logo, la scelta del tipo di edizione da presentare (diplomatica-interpretativa), le modalità di visualizzazione (immagine-testo e

⁴⁶ Unicode è un sistema di codifica che assegna un numero univoco (codepoint) ad ogni carattere usato per la scrittura di testi, in maniera indipendente dalla lingua, dalla piattaforma informatica e dal programma utilizzato.

lettura-testo) e altri elementi cardine per la composizione di EVT 2. Oltre a queste implementazioni basilari, sono stati introdotti anche nuovi valori per quanto riguarda l'identificazione degli inserimenti, il collegamento alla cartella contenente le immagini del manoscritto e la calibrazione dello strumento zoom.

Dopo aver illustrato gli aspetti tecnici inerenti all'elaborazione della visualizzazione del documento digitale, ora ne metteremo in evidenza l'aspetto pratico, descrivendo le modifiche attuate all'interno del file CSS, e mostrando i risultati, così ottenuti, nella presentazione in EVT 2. Tramite l'inserimento dei fogli di stile è stato possibile raffigurare le caratteristiche di alcuni elementi marcati nel file XML e, soprattutto, realizzare un'edizione digitale dinamica che offre numerosi strumenti utili a fini non solo didattici, ma anche divulgativi.

La visualizzazione di default dell'edizione digitale prevede l'immagine della pagina del manoscritto a sinistra, mentre la trascrizione del testo è riportata nella sezione di destra.

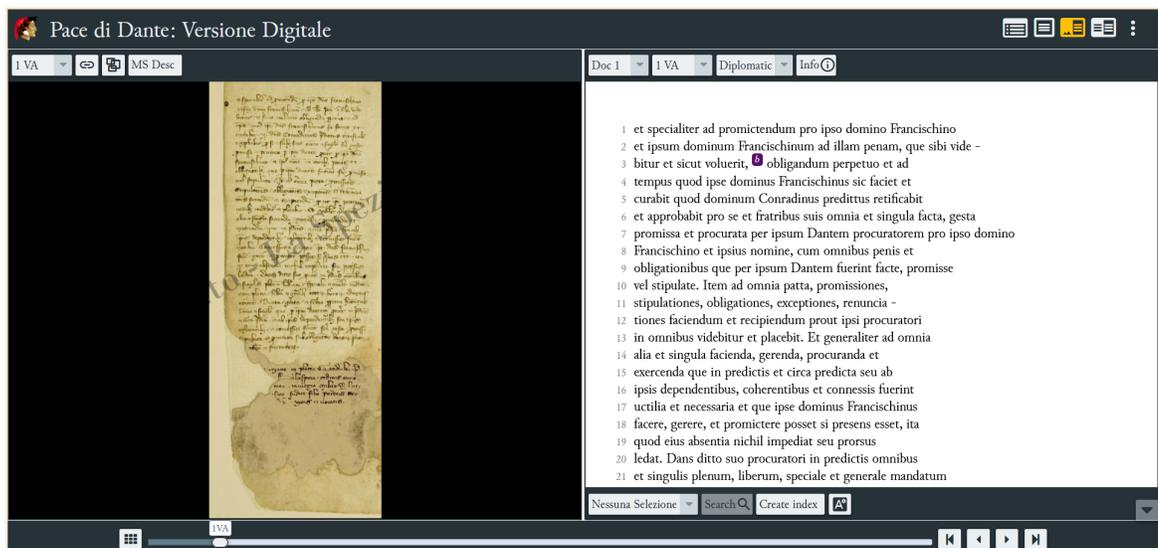


Figura 7: Home page dell'edizione diplomatica-interpretativa in EVT 2

Al di sotto dell'intestazione della pagina, dove troviamo il logo e il titolo dell'edizione, sono presenti dei selettori che mettono a disposizione dell'utente la navigazione sia per pagine, che per documenti. Come già accennato (nr.cap), si è dovuto obbligatoriamente optare per soluzioni anticonvenzionali allo scopo di permettere la simultanea suddivisione tra pagine e documenti. In particolare, mettiamo in risalto due di queste tipologie di espedienti. All'interno della prima pagina, nominata 1RB, le prime 10 righe appartengono al documento 2, mentre le linee restanti riguardano il documento 1. Per

evidenziare questa loro suddivisione si è deciso di marcare il testo delle prime 10 righe con un elemento `<seg type="copy-of">` e di attribuirgli, attraverso le regole CSS, una variazione cromatica differente dal testo sottostante. Inoltre, per rendere chiara la diversa composizione dei documenti 1 e 2, è stata inserita una nota a inizio pagina.

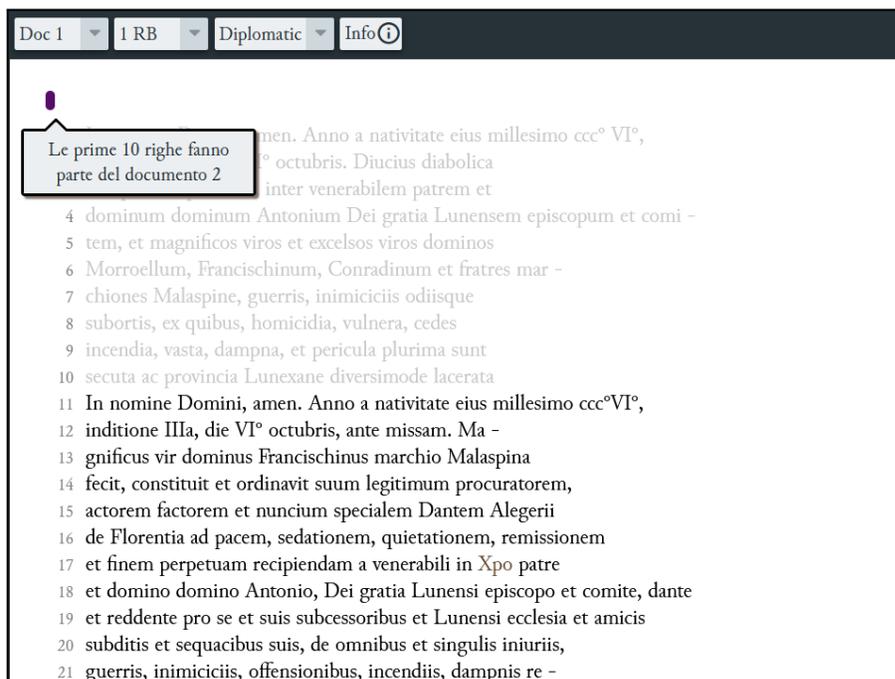


Figura 8: Visualizzazione trascrizione del testo della prima pagina 1RB

Un'altro caso particolare riguarda la pagina 6RB, contenente la fine del documento 3 e l'inizio del documento 4. Per non incorrere in sovrapposizioni di gerarchie e permettere la suddivisione dell'edizione, sia in pagine che in documenti, si è optato per ripartire la pagina in due parti nominate 6RBI e 6RBF. La prima contiene la fine del documento 3 e l'inizio del documento 4 che avrà una colorazione sbiadita attribuita dalle regole CSS. La seconda, viceversa, avrà l'inizio del documento 4 e le linee soprastanti di un colore più spento. La pagina 6RBI apparterrà quindi al documento 3, mentre la pagina 6RBF sarà parte del documento 4. Le parti di testo sbiadite sono state quindi sfruttate per evidenziare la non appartenenza della pagina al documento preso in selezione.

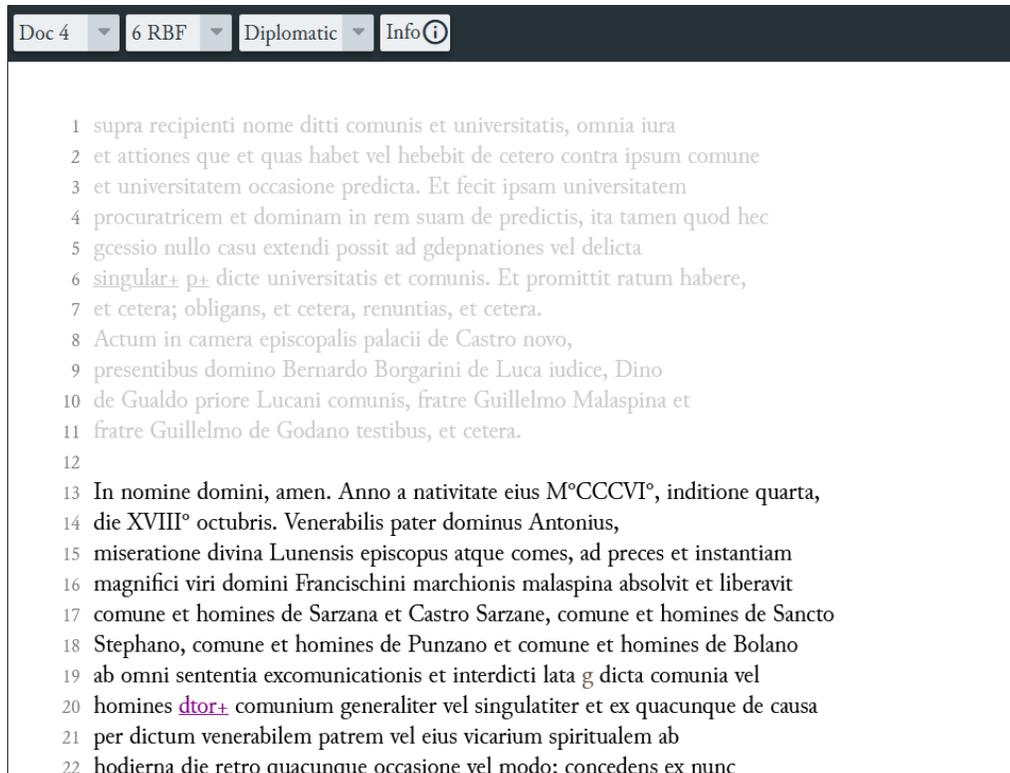


Figura 9: Visualizzazione trascrizione del testo della pagina 6RBF

Come possiamo vedere dall'immagine (nr figura), oltre ai selettori che permettono di navigare tra le pagine e tra i documenti, ci sono anche altri due elementi <button>. Uno consente di variare il tipo di edizione, da diplomatica a interpretativa mentre, l'altro, fa apparire il regesto relativo al documento selezionato.

Tramite il livello di edizione interpretativa vengono messi in risalto, trasformandoli in corsivo, gli scioglimenti di abbreviazione marcati con l'elemento <ex>, figlio dell'elemento <expan>. Gli elementi <choice>, invece, sono resi visibili da un collegamento ipertestuale che, se cliccato, fa comparire un pop-up con su scritto l'espansione relativa alla parola selezionata.

La pagina fornisce la possibilità agli utenti di confrontare il testo latino con l'immagine del manoscritto, scorrendolo riga per riga grazie al collegamento testo-immagine. L'elaborazione di quest'ultimo è stata sviluppata all'interno del file XML, dove abbiamo collegato biunivocamente gli elementi <lb/>, corrispondenti alle righe testo, con le varie porzioni dell'immagine rappresentate dalle <zone>.

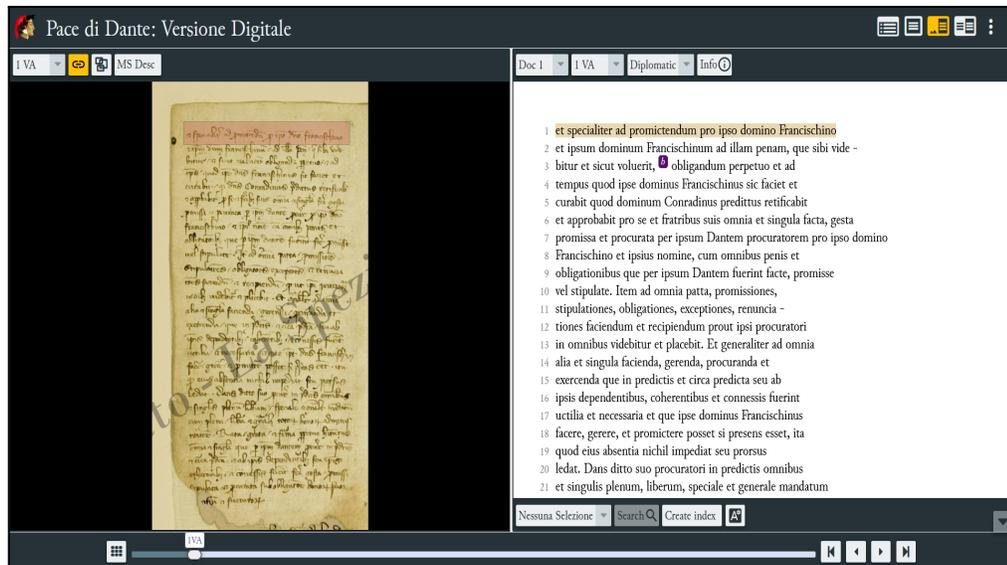


Figura 10: Collegamento testo-immagine in EVT

Bisogna fare un appunto anche per quanto riguarda la rappresentazione degli elementi `<lb/>`, i quali, in alcuni casi, spezzano le parole nell'andare a capo. Per dare un'interpretazione a questo fenomeno, si è aggiunta una regola CSS che fa stampare un trattino prima dell'andare a capo, così da raffigurare chiaramente la frammentazione della parola.

Il testo racchiuso dalla sezione di destra è munito di note che, se cliccate, aprono un pop-up con suscritta la spiegazione di un certo fenomeno testuale.

In fondo alla sezione di destra, troviamo un selettore, il quale permette di evidenziare le named entities marcate precedentemente nel documento XML. In particolare, grazie a questa funzionalità possiamo mettere in risalto le integrazioni, le lacune e le date. Attraverso l'inserimento di regole CSS, inoltre, è stato possibile differenziare le integrazioni, marcate con l'elemento `<supplied>`, tra quelle che sostituiscono le porzioni di testo mancanti a causa del gusto al supporto e quelle aggiunte a causa dell'omissione di testo nel manoscritto.

Le lacune, marcate dall'elemento `<gap>`, sono state rappresentate attraverso tre puntini e delle parentesi quadre che ne rispecchiano la posizione all'interno del manoscritto: se i tre puntini sono seguiti da una parentesi quadra chiusa vuol dire che la lacuna è a sinistra; se sono preceduti dall'apertura della parentesi quadra vuol dire che la lacuna è a

destra; se, infine, risultano interni ai segni di apertura e chiusura della quadra vuol dire che la mancanza di testo si trova al centro.

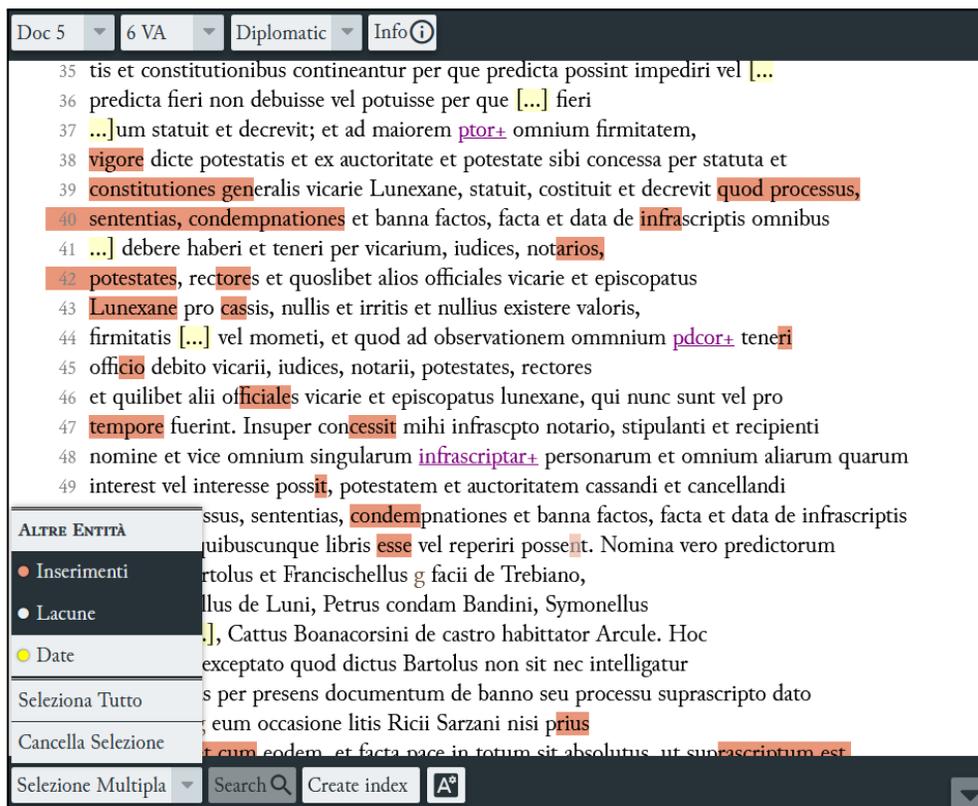


Figura 11: Selettore named entities in EVT2

4.6 Confronto con altri progetti

Lo sviluppo del documento digitale delle pagine appartenenti alla Pace di Dante e le sue relative funzionalità integrate, sono state elaborate analizzando le diverse soluzioni adottate nella realizzazione di altre edizioni digitali basate sulla prima e la seconda versione del software EVT. Una di queste è l'edizione digitale, di tipo diplomatica e interpretativa, del Vercelli Book⁴⁷, prodotta sotto la guida dal professor Roberto Rosselli Del Turco presso l'Università di Torino. Essa è stata adattata per la visualizzazione in EVT 2, rivelandosi un'utile fonte di riscontro, soprattutto per la configurazione del software di presentazione, situata all'interno del file config.json.

⁴⁷ The Digital Vercelli Book: <http://vbd.humnet.unipi.it>.

Traendo spunto dal Vercelli Book⁴⁸, sono state modificate alcune strutture interne al file XML, le quali ci hanno permesso di integrare differenti funzionalità, come ad esempio la rappresentazione ed evidenziazione delle lacune e la suddivisione concomitante tra pagine e documenti. Infine, grazie all'integrazione di regole CSS e una personalizzazione dei valori contenuti nella configurazione di EVT 2, si sono riuscite a mettere in evidenza tutti gli interventi editoriali derivati dall'edizione del Codice Diplomatico dantesco del 2016.

Per quanto riguarda l'edizione critica, compiuta durante il progetto di tirocinio, si è preso in paragone l'edizione digitale del Codice Pelavicino⁴⁹, disponibile su Github. Il prodotto così ottenuto è però risultato differente in molti aspetti, rispetto all'edizione presa come modello. Questo perché la visualizzazione dell'edizione critica è stata elaborata per il software EVT1, comportando una maggior flessibilità nella personalizzazione attraverso le regole di trasformazione XSLT.

4.7 Possibili sviluppi

Il processo di implementazione delle nuove funzionalità di EVT è derivato da sviluppi adottati da alcuni studenti durante tirocini curriculari e tesi triennali presso il corso di laurea di Informatica Umanistica dell'Università di Pisa. Per la maggior parte dei casi, gli studenti sono stati in grado di trarre profitto dal lavoro svolto, arricchendo in maniera considerevole le proprie competenze e fornendo un contributo significativo allo sviluppo di EVT. La difficoltà di questo lavoro sta nel fatto che qualunque nuova caratteristica deve essere integrata ed adattata con quelle preesistenti, richiedendo quindi tempo e pazienza. In più, è presente la problematica legata al fatto che la verifica dei conflitti tra i costituenti di EVT è possibile solo dopo aver inserito il codice di sviluppo costituente la funzionalità aggiunta.

Dopo aver esplicitato la premessa sulle modalità di implementazione di EVT, è ora opportuno addentrarci nello specifico, mettendo in risalto i possibili sviluppi futuri inerenti al progetto finale presentato in questa relazione.

⁴⁸ Edizione digitale del manoscritto redatto nell'Inghilterra meridionale alla fine del X secolo, importante per le ricerche sulla lingua e letteratura anglosassone. (Archivio e Biblioteca Capitolare di Vercelli CXVII).

⁴⁹ Il Codice Pelavicino è un manoscritto del XIII secolo conservato presso l'Archivio Capitolare Lunense (Sarzana) che si compone di 426 carte e contiene in copia documenti dall'anno 900 all'anno 1289. Edizione digitale del Codice Pelavicino: <http://pelavicino.labcd.unipi.it>.

Iniziamo col parlare della possibilità di inserire una traduzione italiana dei documenti. L'integrazione di questa funzionalità sarebbe possibile configurando il software di presentazione EVT, in modo tale da garantire una corretta visualizzazione della traduzione e selezionarla proprio come se fosse un livello di edizione a sé stante. La presenza di una versione italiana dei documenti appartenenti alla Pace di Dante accrescerebbe l'accessibilità dell'opera stessa, permettendo, anche a chi non ha conoscenza del latino, di comprendere appieno il contenuto testuale dei vari atti. Da un punto di vista quantitativo, il progetto potrebbe ampliarsi ulteriormente tramite l'inserimento di altri atti appartenenti al manoscritto. Questo sarebbe possibile solamente proseguendo la collaborazione con l'Archivio storico di La Spezia, dove sono contenuto il complesso documentario dantesco. Inoltre, la mia edizione potrebbe servire come modello di riferimento o, per meglio dire, punto di partenza per successive edizioni del manoscritto, da parte di tirocinanti o studiosi, così da alleggerirne il peso della codifica e la parte legata alla progettazione concettuale. Infine, un'altra possibile miglioria del progetto potrebbe corrispondere all'integrazione di un'edizione critica sviluppata in EVT 2. L'apparato critico dovrebbe comprendere tutte le variazioni testuali dei singoli atti, con informazioni relative alle loro edizioni di appartenenza e ai protagonisti che le hanno curate. L'elaborazione di questo tipo di edizione richiederebbe un gran lavoro di studio e approfondimento dei vari testi, ma d'altro canto, garantirebbe una completa versione dei singoli atti del documento ponendo in parallelo, su più livelli, tutte le diverse variazioni testuali appartenenti alle diverse edizioni. Mettendo, infine, in evidenza gli interventi editoriali dei curatori di ciascuna edizione, si riuscirebbe ad accrescere notevolmente l'intelligenza artificiale del documento stesso.

5 Conclusioni

Il progetto illustrato si presenta come un utile strumento di studio per l'epistolario dantesco, mettendo insieme le immagini del manoscritto della Pace di Dante e la corrispondente trascrizione in lingua latina. L'edizione digitale è stata integrata dagli interventi editoriali presenti nell'edizione del Codice diplomatico dantesco del 2016, i quali, oltre a semplificare la comprensione del testo, danno anche informazioni utili circa il contesto storico in cui sono state composte le lettere. La possibilità di sviluppare altre edizioni del medesimo manoscritto, garantirebbe una maggior conservazione ed accessibilità del patrimonio documentario dantesco, fornendo strumenti utili per studi meticolosi.

La creazione dell'edizione digitale dei documenti della Pace di Dante mi ha consentito di scoprire quanto sia importante l'ambito dell'informatica umanistica nel conciliare competenze informatiche con le conoscenze umanistiche e, soprattutto, quanto la tecnologia possa facilitare la divulgazione del patrimonio documentario che spesso giace nascosto negli archivi, rendendolo accessibile e condivisibile a tutti.

6 Bibliografia

Alighieri Dante, *Codice diplomatico dantesco: I documenti della vita e della famiglia di Dante Alighieri*, a cura di Guido Biagi e Lando Passerini, 1895.

Alighieri Dante, *Codice diplomatico dantesco*, a cura di Renato Piattoli, 1950.

Alighieri Dante, *Nuova edizione commentata delle opere di Dante*, Vol. 7\3: Opere di dubbia attribuzione e altri documenti danteschi: Codice diplomatico dantesco, Salerno editore, 2016, a cura di De Robertis Teresa, Milani Giuliano, Regnicoli Laura e Zamponi Stefano.

Ciotti Fabio, *Breve introduzione alla Text Encoding Initiative*, in Biblioteca Italiana, 1 dicembre 2003.

Ciotti Fabio, *Cosa è la codifica informatica dei testi*, in AA.VV., Atti del Convegno Umanesimo & Informatica (Trento 24-25 maggio 1996), a cura di Daniela Gruber e Patrick Pauletto.

Ciotti Fabio, Roncaglia Gino, *Il mondo digitale. Introduzione ai nuovi media*, Roma-Bari, Editori Laterza, 2001.

Gigliozzi Giuseppe, *Il testo e il computer. Manuale di informatica per studi letterari*, Milano, Bruno Mondadori, 1997.

Shepherd Devan, XML. *Guida completa, trad. it.* staff Apogeo, Milano, Apogeo, 2002.

Theodor Holm Nelson, *Literary Machines 90.1*, Padova, Franco Muzzio Editore, 1992.

7 Sitografia

Adobe Photoshop: <http://www.adobe.com/it/products/photoshop.html>.

Codice Pelavicino digitale: <https://pelavicino.labcd.unipi.it/>.

EVT: <https://github.com/evt-project/evt-viewer/>.

HTML (HyperText Markup Language): <http://www.w3.org/HTML/>.

JavaScript: <https://www.javascript.com/>.

jQuery: <https://jquery.com/>.

SGML: <https://www.iso.org/standard/16387.html/>.

Text Encoding Initiative (TEI): <http://www.tei-c.org/>

TEI Consortium, TEI P5: Guidelines for Electronic Text Encoding and Interchange. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>.

Unicode Consortium: <http://www.unicode.org>.

Vercelli Book digitale: <http://vbd.humnet.unipi.it/>.

W3C-CSS (Cascading Style Sheets): <https://www.w3.org/TR/css/>

W3C-XSL Transformation: <https://www.w3.org/TR/xslt/>.

XML: <https://www.w3.org/XML/>.